

Face Recognition Based On Gabor Local Feature and Convolutional Neural Network

Weimeng Qin ^a, Lie Wang ^b and Wen Luo ^c

School of Computer and Electronics Information Engineering, Guangxi University, Nanning
530004, China

^a812608001@qq.com, ^blwang@gxu.edu.cn, ^c120758954@qq.com

Keywords: Gabor features, Convolutional neural network, Face recognition.

Abstract. Since the distribution of kernel function of Gabor transform and the Two-Dimensional receptive field profiles of mammalian simple cells in the primary visual cortex is very similar, and has the direction selectivity and good spatial locality, so the acquisition of spatial scale information of multiple directions and local structure features in the local regions of images provide a more effective method. The method is based on the Gabor transform and fused the convolutional neural network of the powerful learning ability. The local features of the Gabor transform is used as the input of the neural network, the neural network is used to classify the samples. In ORL face database, the experimental results show that the face recognition methods based on Gabor local features and convolutional neural network in the same conditions obtain higher face recognition rate and the robustness.

1. Introduction

With the development of information technology, the technology of face recognition is widely used in all aspects of life, which has become a hot topic in the field of artificial intelligence and pattern recognition because of its great application prospect in the fields of public security, customs, authentication, medical science and so on. At present, due to the influence of face pose and illumination factors, it brings great difficulties to face feature extraction, and the face recognition rate needs to be further improved. Therefore, how to efficiently extract facial features has been a difficult problem in the field of face recognition^[1].

The results show that the kernel function distribution of Gabor transform is similar to the Two-Dimensional receptive field profile of mammalian visual cortical cells, which has excellent orientation and frequency selectivity. Lades, M, et al used gabor transform in face recognition research for the first time^[2]. The gabor feature has good robustness to the local change of face caused by pose and external factors, and it is also simple. On the other hand, the algorithm is complex and the computation is very huge and time-consuming. The method proposed in this paper avoids the elasticity of the high computational complexity of the matching algorithm, and the memory capabilities can greatly improve the fault tolerance of the algorithm by the neural network's unique learning, which can improve the recognition rate, reduce the running time and the operation cost.

2. Two-Dimensional Gabor transform

In 1946, Dennis Gabor proposed a localized time window function to achieve signal local information in the Fourier transform as the Fourier transform can not analysis the local time-frequency at the same time^[3], which was known as the "Window" of Fourier transform(It is also called the short-time Fourier transform, STFT). The later generations call it the Gabor transform, which plays a very important role in the non-stationary signal analysis.

It provides the possibility for the precise localization feature extraction of Two-Dimensional images because of the highly accurate localization information obtained by the Gabor transform. In 1985, Daugman et al applied the Gabor transform to the two-dimensional space at the first time, which transformed the Gabor wavelet into two dimensional form and constructed the two-

dimensional Gabor wavelet^[4]. According to the known characteristics of the visual system cells, they put forward that the simple cell activities on human visual cortex can be simulated by the two-dimensional Gabor transform, which used a set of different scales and directions of the two-dimensional Gabor filter to simulate the characteristics of the visual cortex cells, and the result of filtering two dimensional signals through this group of two-dimensional Gabor filter can be took as the responses of visual cells stimulated by external signals.

The kernel functions of the two-dimensional Gabor transform can be formulated as:

$$\psi(x, y, \omega_0, \theta) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x_0^2+y_0^2)}{2\sigma^2}\right) (\exp(j\omega_0 x_0) - \exp(\frac{-\omega_0\sigma^2}{2})) \quad (1)$$

$$x_0 = x\cos\theta + y\sin\theta \quad (2)$$

$$y_0 = -x\sin\theta + y\cos\theta \quad (3)$$

In the above formulas, x and y determine the location of the pixels in the spatial domain together, θ represents the direction of the Gabor filter, ω represents center frequency, σ^2 represents the variance of the Gaussian function. $\exp(-\omega_0\sigma^2/2)$ represents DC component, which can reduce the interference of the Gabor filter from the absolute gray value of the image so that the Gabor filter is insensitive to the image with different illumination. $\exp(-(x_0^2+y_0^2)/2\sigma^2)$ represents Gaussian function which can limit the range of $\exp(j\omega_0 x_0)$ so as to achieve the purpose of local optimization.

By the formal definition of the kernel function of the two-dimensional Gabor transform, it can be concluded that the two-dimensional Gabor filter not only has good discrimination ability in the spatial domain, but also inherits this characteristic in the frequency domain. By choosing different θ , it has good directional selectivity in the spatial domain; By adjusting the value of ω_0 , it has better frequency selectivity in frequency domain. Typically, when the value of ω_0 and θ is determined, it is equivalent to constructing a Gabor filter. In the experiment process, a different direction and different scale Gabor filter group is composed of a plurality of Gabor filter through choosing different ω_0 and θ in order to obtain better and more accurate localization signal. In this paper, the Gabor filter banks were chosen which were used in the face recognition experiments usually. The Gabor filter bank consist of 40 filters that each one had five different scales and eight different directions^[6]. The specific parameters were as follows:

$$\omega_0 = 2^{-\left(\frac{f+2}{2}\right)}\pi \quad (4)$$

$$\theta = \frac{\pi}{8}r \quad (5)$$

$$f \in \{0,1,2,3,4\}, r \in \{0,1,2, \dots, 7\}.$$

3. Convolution Neural Network

The convolutional neural network is mainly divided into forward and backward propagation, and the convolution layer and sampling layer are alternately carried out. The sampling layer after the convolution layer is used to construct the invariance of space and structure and reduce the running time.

The forward propagation is calculated from the input parameter to the output. The output of the first layer is the input of the current layer, and then the output of the current layer is passed by layer by activation function. The output of the current layer can be expressed as:

$$x^l = f(W^l x^{l-1} + b^l) \quad (6)$$

In the above formula, l represents layer number, W express the weight, f is the activation function. The hyperbolic tangent function is used in this paper.

Back propagation is a combination of forward propagation calculation, doing error computation by using a sample label, considering the square loss function. For the multi class problems which contains c categories and N multi classification training samples, the error function can be expressed as:

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (7)$$

In the above formula, t_k^n Represents the n dimension of the corresponding label in the k sample. y_k^n represents the n output of the corresponding network output in the k sample. The back

propagation can update the convolution layer, while the feature mapping of the upper layer and a trained kernel can perform convolution operations, and the operation result can form the feature mapping after the activation function is formed. Therefore, each output is related to the convolution of several feature mappings in the previous layer. The form of the convolution layer is usually as follows:

$$x_j^l = f\left(\left|\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right|\right) \quad (8)$$

In the above formula, l represents the layer number, k represents the convolution kernel, M_j represents a choice of input features, b is a bias.

The down sampling reduces the size of feature mappings. Assume that the size of the sampling operator is $n \times n$, the size of the feature map becomes $1/n$ of the original feature size after a sub sampling. The general form of down sampling is as follows:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l) \quad (9)$$

In the above formula, `down()` represents a lower sampling function.

The training process of the convolution neural network is the learning process: To obtain local features as samples by using Gabor transform, the neural network changes the connection weights of the network under the stimulation of input samples so that the output of the network can approach the desired output continuously. The essence of the learning is to adjust the weights of each connection dynamically so that the output can achieve the desired goal. The learning rule is a set of adjustment rules based on the connection weights of each neuron in the process of learning. The overall flow chart is shown in Figure 1:

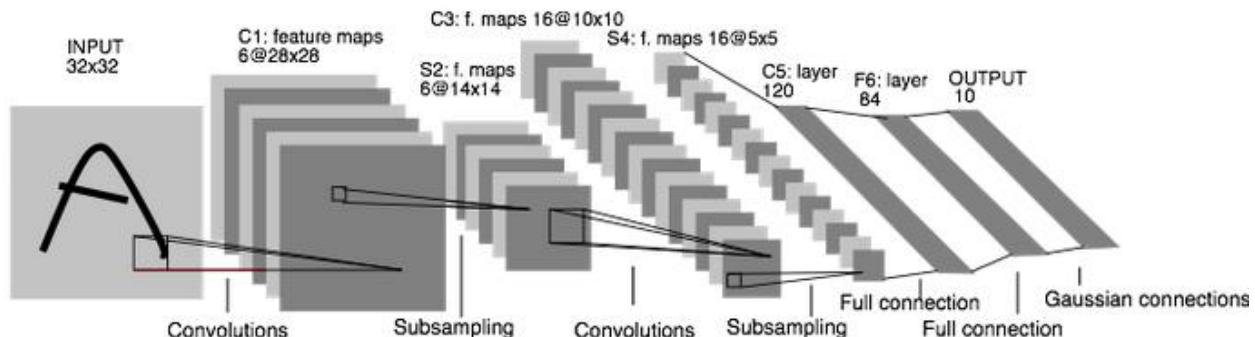


Fig. 1 Convolution network structure diagram

4. The Result Analysis Of Face Recognition Experiment

4.1 The face recognition data set

In this paper, the three authoritative datasets of face recognition: ORL, YALE and CMU_PIE face databases were used.

ORL face database, which is collected by AT&T Laboratory of, University of Cambridge, is one of the simplest and most efficient face databases in face recognition research at present. It consists of 40 people of different ages, sexes and races, each of whom has 10 facial images (including changes in facial expressions, gestures, and occlusion of face ornaments), consisting of 400 pictures in total [8]. YALE face database is created by the computer vision and control center of Yale University. The database has participated in 15 volunteers and a total of 165 pictures which were taken every 11 gray-scale images in different light conditions and different expression changes. The CMU_PIE face database was created in by Carnegie Mellon University in November 2000, which were mainly for pose, illumination and expression. It contains 68 people, each of it has 13 pose conditions, 43 light conditions and 4 facial expressions, which has 40000 photos. Thanks to the huge picture size, it is suitable for the neural networks training.

4.2 Experimental environment configuration and parameter setting

This article uses Matlab 2009a on Windows7 as the experimental environment, the CPU of the computer is Intel i5 processor, 3.20GHz clocked, 8GB memory and the graphics card is Nvidia GTX750.

Face alignment 2D alignment according to the position of the eye and nose. Set the convolutional network training parameters: The size of convolution kernel in two convolution layers is 5×5 , use the maximum pool in the two sampling layers, the sample size is 2×2 , the sigmoid function uses activation function.

4.3 Database simulation experiment

Before the experiment, the CMU_PIE face database was used as training sample, and the ORL database and YALE database were used as test samples to test the recognition effect of different algorithms under the interference of expression, gesture and occlusion. The resolution of the standard test image after processing is 130×96 .



Fig. 2 Face recognition test case diagram in ORL

In this paper, the six popular methods are compared to verify the effectiveness and accuracy of the proposed algorithm. The six methods are: GNN algorithm^[9], LBP algorithm^[10], FF algorithm^[11], PCA algorithm^[12], ICA algorithm^[13] and TPFRS algorithm^[14].

Using different methods, the system identification rate is different. Table 1 shows the recognition rate of the proposed method and the recognition rate of the above six methods under the ORL database:

Table1 Contrast and compare different algorithms

Methods	Recognition rate (%)		Execution efficiency (s)	
	ORL	YALE	ORL	YALE
GNN	96.75	88.48	0.53	0.47
LBP	94.00	84.24	0.48	0.33
FF	94.75	85.45	0.39	0.41
PCA	87.00	72.12	0.21	0.29
ICA	85.25	64.85	0.41	0.21
TPFRS	98.50	90.30	2.57	2.34
The Proposed Method	99.75	95.76	0.47	0.39

Table 1 shows the accuracy and efficiency of the algorithm in comparison with GNN, LBP, FF, PCA, ICA, and TPFRS on six standard databases on two standard databases. From table 1, the proposed algorithm achieves the best recognition accuracy on two databases. Compared with the TPFRS method using the traditional sparse expression model, the proposed method improves the recognition accuracy by 1.26% and 6.04% on the two databases of ORL and YALE, with an average improvement of 3.65%. In terms of execution efficiency, the average time of recognizing a face image by using proposed method in database is 0.53s and 0.47s. Compared with the TPFRS algorithm, the recognition time is 2.57s and 2.34s, and the recognition efficiency is improved by about 5 times. The algorithm proposed in this paper need to extract local features through Gabor, and then classify them by trained neural networks. That is why it is time-consuming compared with other algorithms. But on the other hand, the algorithm proposed in this paper has high recognition accuracy and good running time in the comprehensive recognition rate and the execution efficiency.

4.4 Influence of training samples number on recognition accuracy

This paper also tested the effect of the number of training samples on the overall recognition performance, figure 3 shows the algorithm compared to GNN, FF and TPFRS three kinds of algorithm in ORL database and YALE database with the impact of changes in the number of training samples for recognition accuracy.

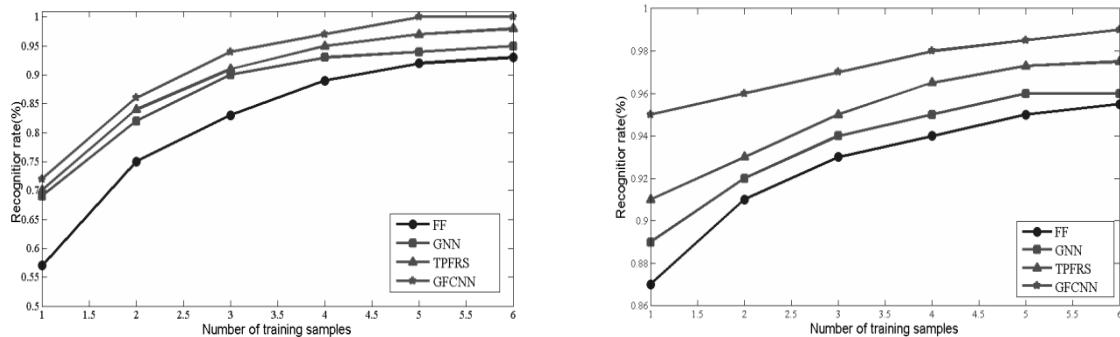


Fig. 3 The change of identification accuracy with the training sample number on the ORL and YALE database

As shown in Figure 3, the algorithm changes the recognition accuracy in the ORL and YALE databases as the number of training samples changes. In this paper, FF, GNN and TPFRS algorithms are chosen as benchmark. Obviously, as the number of training samples increases, the recognition accuracy of each method is increasing. In contrast, the algorithm proposed in this paper has the highest recognition accuracy and is obviously better than the other algorithms.

5. Conclusion

Gabor features are widely used in visual information understanding, which provides a great help by selectivity direction and good spatial locality for the research and application of face recognition. And the convolutional neural network is effective to solve the computational complexity problem of the traditional elastic graph matching algorithm, without elastic rough matching and exact matching, only through training complete classification. The real-time performance of the traditional algorithm is greatly improved in the identification, and the biological characteristics of the algorithm is preserved through the Gabor transform. The learning and classification ability of the convolutional neural network is better than traditional elastic graph matching at the same time. Compared with some popular algorithms, the experimental results show that the proposed method is superior to other face recognition algorithms in ORL and YALE database. Therefore, the algorithm presented in this paper is a face recognition method with high recognition accuracy and robustness. The follow-up will focus on how to apply the algorithm proposed in this paper to the recognition of face images in video sequences.

References

- [1]. LI A, SHAN S, GAO W. Coupled bias-variance tradeoff for cross-pose face recognition[J].*Image Processing, IEEE Transactions*, 2012, 21(1):305–315.
- [2]. LADES M, VORBRUGGEN J C, BUHMAN J, et al. Distortion invariant object recognition in the dynamic link architecture[J].*Computers, IEEE Transactions*, 1993, 42(3):300–311.
- [3]. GABOR D. Theory of communication[J].*Journal of the institute of Electrical Engineers*, 1946, 93(26):429–457.
- [4]. DAUGMAN J G. Complete discrete 2D Gabor transforms by neural networks for image analysis and compression [J].*IEEE Transactions on Acoustic, Speech and Signal Processing*, 1988, 36(7):1169–1179.
- [5]. LEE T S. Image representation using 2D Gabor wavelets[J].*IEEE Trans Pattern Analysis and Machine Intelligence*, 1996, 18(10):959–971
- [6]. LIU C. Gabor-based kernel PCA with fractional power polynomial models for face recognition[J].*IEEE Trans Pattern Analysis and Machine Intelligence*, 2004, 26(5):572–581.
- [7]. SYAFEEZA A R, KHALIL H M, LIEW S, et al. Convolutional neural network for face recognition with pose and Illumination Variation[J].*International Journal of Engineering & Technology*, 2014, 6(1):44–57.

- [8]. DUDA R, HART P, STORK D. Pattern classification[M].2nd ed. Beijing: Beijing Publishing House, 2001.
- [9]. YANG A Y, ZI H Z, GANESH B A, et al. Fast-minimization algorithms for robust face recognition[J].*IEEE Trans Pattern on Image Processing*, 2013,22(8):3234–3236.
- [10]. AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: Application to face recognition[J].*IEEE Trans Pattern Analysis and Machine Intelligence*, 2006,28(12):2037–2041.
- [11]. BELHUMEUR P N, HESPAÑA J P, KRIEGMAN D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J].*IEEE Trans Pattern Analysis and Machine Intelligence*, 1997,19(7):711–720.
- [12]. TURK M, PENTLAND A. Eigenfaces for recognition[J].*Journal of Cognitive Neuroscience*, 2010,3(1):71–86.
- [13]. BARTLETT M S, MOVELLAN J R, SEJNOWSKI T J. Face recognition by independent component analysis[J].*IEEE Transactions on Neuro Network*, 2002,13(6):1450–1464.
- [14]. WAGNER A, WRIGHT J, GANESH A, et al. Toward a practical face recognition system: Robust alignment and illumination by sparse representation[J].*IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012,34(2):372–386.